

Status Report on CERS Data Exchange and Data Dictionary Issues

Cal/EPA Unified Program November 29, 2010

The Cal/EPA Unified Program staff prepared this document for the November 30th, 2010 Data Standards Steering Committee Meeting to summarize recent efforts and current/upcoming issues related to CERS data exchange planning and data dictionary issues.

CERS Data Standard: XML versus “Flat File” (Delimited) Formats

Issue	The DSC’s previous decision to support both XML and “flat file” data formats for loading data into CERS will involve significant technical effort. The decision should be revisited to confirm: 1) the business needs/drivers, 2) at what point in the project lifecycle must flat file be supported in CERS2, and 3) should flat file support be ongoing or phased out at some point in the future.
Considerations/Impacts	XML is the modern standard for electronic data exchange, but tends to be more technically complex to implement than flat file data. Unlike flat file formats, XML is especially adept at organizing hierarchical data which is commonly found in the Unified Program (e.g., a facility’s multiple UST tanks or Recyclable Materials data records), as well as allowing more flexibility when data elements are added/removed. Supporting flat file uploads into CERS, especially for multiple facilities at a time, will add considerable short term technical effort, as well as ongoing support costs as data standards change over time. Cal/EPA staff has asked whether they may be more cost-effective ways in the short or long term to migrate all UPAs to XML-based uploads (e.g., grants). Note: Downloads of data from CERS2 in flat file data for a specific program element is not a technical problem.
Current Status	During discussion of this issue at recent CERS Change Management Committee and Regulator User Group meetings, participants identified that flat file upload of data is of equal if not more importance to businesses than regulators. In particular, managing chemical inventory data via flat file upload. Recent analysis by Cal/EPA Unified Program Technical Services staff has indicated that flat file uploads for a single program element for a single facility will generally not be that difficult. However, support for uploading multiple program elements for one or multiple facilities may begin to approach the same complexity as XML data formats.
Action(s) Needed	Confirm the need for flat file upload support and when it must be available in CERS2. Consider whether some future deadline should be agreed upon for phasing out of flat file support.

Locally-Collected Information

Issue	<p>CERS2 will support manual web forms as well as machine-to-machine, electronic data transfer (EDT) submittal between businesses, regulators, and CERS. CERS' EDT scenarios are quite complex, involving three parties and multiple paths (see Appendix A). Successful and cost-effective EDT requires data exchanges to be fully and rigorously defined in a technical document ("exchange schema") so programming can replace the decision-making normally performed by humans using manual forms. Having an ad hoc body of locally collected fields used by some UPAs and not others precludes Cal/EPA from publishing a single "exchange schema" that can be used statewide by all businesses, regulators, and data management software vendors. Cal/EPA proposed three options to this problem in an issue paper provided to the CERS Regulator User Group.</p>
Considerations/Impacts	<p>Below are the options provided to the User Group—the descriptions have been slightly modified for brevity and to reflect the Group's discussions.</p> <p>Option 1: CERS2 Does NOT Support ANY Locally Collected Fields</p> <p>The CERS2 web pages and electronic reporting would not include any locally collected fields—only fields described in the Title 27 Data Dictionary (or required for general CERS operation). UPAs would need to acquire (and store) any locally collected fields from the business outside of CERS. UPAs could provide instructions and links to their businesses to complete local forms on the CERS "Local" tab. Multi-facility/multi-jurisdictional businesses would be forced to discover/provide locally collected information from each UPA.</p> <p>Option 2: CERS2 Supports a Collection of Local Fields Approved for Statewide Use</p> <p>UPAs and Cal/EPA agree upon a small selection of fields required of all businesses statewide in CERS2 web forms and data exchanges. These fields would be included in a future rulemaking for addition to the Title 27 Data Dictionary. All of these formerly local fields would be included in the single statewide "exchange schema", reducing the number of special cases multi-facility/multi-jurisdictional businesses would need to handle.</p> <p>Option 3: CERS2 Supports a Moderate Collection of pre-defined Local Fields</p> <p>In CERS2 UPAs would select from an approved master list of pre-defined, locally collected fields (never to exceed ~100-150 fields due to technical limitations). The UPA's businesses would be prompted to complete these fields in the CERS2 web pages, and encouraged to provide this data in EDT submittals. All pre-defined local fields would be included in the statewide "exchange schema" as optional fields from a statewide basis. CERS2 would NOT enforce/validate/audit for values in these fields for a given UPA/business submission, requiring UPAs to ensure local data is provided (or contact the business to solicit these extra fields). Additions to this master list of local fields would be rigorously reviewed and approved by a state level change management group.</p> <p>Option 3 involves considerably more technical effort for Cal/EPA. Either Option 2 or 3 involves incremental technical effort by any EDT-bound UPAs (or their data software vendors) for each additional field added.</p>
Current Status	<p>After Cal/EPA conducted an initial survey of locally collected fields during Sept.-Oct.2010 resulting in ~100 fields, and a follow-up survey of only business critical fields that resulted in ~20 fields, the User Group voted to eliminate Option 1 (on Nov.9) and adopted Option 3 at their Nov.29th meeting (17 UPAs + 2 state vs. 12 for Option 2).</p>
Action(s) Needed	<p>Review the options proposed by Cal/EPA to the CERS Regulator User Group and either adopt the User Group's recommendation, or provide alternate direction to Cal/EPA. Note: Further study/research will delay EDT standards development beyond the current February/May 2011 specification schedules.</p>

Inspection/Enforcement/Violation Data Exchange

Issue	The current Title 27 Data Dictionary specification for inspection data only includes violation summary data, while the enforcement specification does not include any violation-related data fields. CERS2 EDT will support both inspection and enforcement data using the existing specifications (with some minimal changes). Additionally, Cal/EPA intends to add expanded EDT of inspection and enforcement data to CERS2 that includes detailed violation information for inspections (including a system of violation IDs defined in a Violation Dictionary) and a listing of relevant violations for an enforcement. A number of clean-up and clarification changes are required to support these specifications, and several other suggested changes are strongly recommended by Cal/EPA as well.
Considerations/Impacts	The expanded versions of inspection/enforcement/violation EDT address various statutory requirements not yet fully expressed in Title 27 regulations. Cal/EPA intends to work toward incorporating all of these expanded elements and corrections into a future regulatory update of the Title 27 Data Dictionary.
Current Status	Cal/EPA Unified Program staff is presenting a detailed discussion of this issue to the DSC in the document title "Implementation of CERS2 Electronic Data Transfer (EDT) for Inspection, Enforcement, and Violation Data." This document was emailed to DSC members on Nov.23 and provided at the Nov.30 DSC meeting.
Action(s) Needed	Review the document and either adopt some/all recommended changes to Title 27 Data Dictionary definitions, , or provide alternate direction to Cal/EPA. Note: Further study/research will delay EDT standards development beyond the current February/May 2011 specification schedules.

Document Handling in CERS2

Issue	There are a number of Unified Program forms that are not currently defined in the Title 27 Data Dictionary and will not be included in CERS2. However, a number of documents will need to be uploaded as standalone electronic documents (e.g., PDFs) by businesses to constitute a complete annual submittal (e.g., plot plan, Training Plan, several non-UPCF UST forms). Cal/EPA Unified Program technical staff has identified significant complexities in receiving these document uploads via EDT, requiring various “back-channel” business processes separate from the primary EDT submissions.
Considerations/ Impacts	<p>Cal/EPA has determined that CERS2 will initially include only the standard forms already defined in Title 27 Data Dictionary. This will exclude several forms that were in Unidocs or earlier versions of CERS 1. The UST Financial Responsibility was disabled during Summer 2010, and the UST Response Plan and UST Statement of Designated Operator forms will be similarly disabled in CERS1 in December 2010. CERS1 and CERS2 will support uploading all of these documents as standalone electronic documents.</p> <p>Both Cal/EPA and UPAs will experience increasing storage costs for uploaded documents over time, and will not gain any of the benefits of having the data in these documents stored in database format like the other standard UPCF forms.</p>
Current Status	Cal/EPA Unified Program technical staff are developing the “back-channel” EDT processes and business rules required for document uploads.
Action(s) Needed	Cal/EPA Unified Program recommends the DSC (or a work group(s)) expedite work on standardizing as many Unified Program forms as possible into standard forms (and sets of data fields) so they can be incorporated into CERS as electronic data rather than standalone electronic documents. This will simplify reporting for businesses and reduce costs for UPAs and Cal/EPA.

Request to Make Billing-Related Fields Mandatory in CERS2

Issue	The CERS Regulator User Group at its November 29 th meeting adopted a recommendation for the DSC to adopt several billing-related fields as statewide required fields, rather than as locally collected fields (~28 for, 1 opposed).
Considerations/ Impacts	<p>The fields are listed below. Their exact specifications were not explicitly defined, but would presumably follow similar data field definitions used elsewhere in the Title 27 Data Dictionary.</p> <ul style="list-style-type: none">• Billing Contact Name• Billing Contact email address• Billing Contact Address• Billing Contact City• Billing Contact Zip Code• Billing Contact Phone <p>Apparently many/most UPAs already collect these fields in some fashion from their businesses.</p> <p>Adopting these as statewide-required fields will ensure they are required for all businesses, reducing reporting ambiguity for businesses using EDT and reducing workload on UPAs to confirm these fields are completed (as required for other local fields under Option 3).</p> <p>Cal/EPA Unified Program technical staff welcome any decisions like this that reduce ambiguity and lessen dependence on locally collected fields.</p>
Action(s) Needed	The DSC should consider the recommendation from the CERS Regulator User Group.

Upcoming Additional Title 27 Data Dictionary Modifications

Issue Cal/EPA Unified Program technical staff have been rigorously examining the existing Title 27 Data Dictionary field definitions, the existing Unified Program Consolidated Forms (UPCFs), and comparing them against existing and future CERS business requirements. In addition to the inspection/enforcement/violation changes being brought to the DSC in November 2010, a significant number of other modifications will need to be made to correct errors and inconsistencies in the existing Title 27 Data Dictionary. Cal/EPA also anticipates proposing a number of data dictionary modification recommendations to the DSC that will assist in implementing CERS2/EDT and/or avoid data problems for UPAs and Cal/EPA in the future.

Considerations/Impacts In some cases these recommended changes will be based upon data problems already found in Unidocs/CERS1 data, and in other cases the recommendations will address likely data problems that will likely occur in the future once CERS contains complete submittal data for 100,000+ businesses.

Recommendations are not yet ready to provide to the DSC, but a sampling of errors and/or problems that have been identified include:

- Inconsistent option selections between Title 27 and UPCF forms (error).
- Fields that don't have a type or length explicitly listed in the field definition (error).
- Fields lengths that are too short to meet the identified business requirements (e.g., the current phone numbers definitions (AN 15) can't readily represent extension numbers or international numbers; many person/business name fields are defined as 20-25 AN and probably should be longer).
- Changing fields that are obviously numeric (e.g., fine amount, number of tanks) from alphanumeric fields to numeric fields.
- Addition of fields to handle certain business cases (e.g., addition of the facility's CERS ID to most areas of the data dictionary; addition of a Country field to handle international owner/operator contact information; breaking up the Fire Code Hazard classification field into multiple fields).

Cal/EPA Unified Program staff will need review/approval of the recommended changes by mid-January to meet the CUPA Conference target date communicated to CUPAs in Summer 2011 for the complete field list required for EDT.

Action(s) Needed Cal/EPA recommends the DSC hold a phone conference meeting in the January 10-18, 2010 timeframe to review, discuss, and decide upon the recommended Title 27 data dictionary changes.

Appendix A: CERS2 Data Flows Diagram

